

## The effect of testing on skills learning

Charles B Kromann, Morten L Jensen & Charlotte Ringsted

**OBJECTIVES** In addition to the extrinsic effects of assessment and examinations on students' study habits, testing can have an intrinsic effect on the memory of studied material. Whether this testing effect also applies to skills learning is not known. However, this is especially interesting in view of the need to maximise learning outcomes from costly simulation-based courses. This study was conducted to determine whether testing as the final activity in a skills course increases learning outcome compared with an equal amount of time spent practising the skill.

**METHODS** We carried out a prospective, controlled, randomised, single-blind, post-test-only intervention study, preceded by a similar pre- and post-test pilot study in order to make a power calculation. A total of 140 medical students participating in a mandatory 4-hour in-hospital resuscitation course in the seventh semester were randomised to either the intervention or control group and were invited to participate in an assessment of learning

outcome. The intervention course included 3.5 hours of instruction and training followed by 30 minutes of testing. The control course included 4 hours of instruction and training. Participant learning outcomes were assessed 2 weeks after the course in a simulated scenario using a checklist. Total assessment scores were compared between the two groups.

**RESULTS** Overall, 81 of the 140 students volunteered to participate. Learning outcomes were significantly higher in the intervention group ( $n = 41$ ; mean score 82.8%, 95% confidence interval [CI] 79.4–86.2) compared with the control group ( $n = 40$ ; mean score 73.3%, 95% CI 70.5–76.1) ( $P < 0.001$ ). Effect size was 0.93.

**CONCLUSIONS** Testing as a final activity in a resuscitation skills course for medical students increases learning outcome compared with spending an equal amount of time practising the skills.

**KEYWORDS** clinical competence/\*standards; randomised controlled trial; prospective studies; pilot projects; comparative study [publication type]; resuscitation/\*education; \*education, medical, undergraduate; feedback.

*Medical Education* 2009; **43**: 21–27  
doi:10.1111/j.1365-2923.2008.03245.x

Centre for Clinical Education, Rigshospitalet, Copenhagen University Hospital, Capital Region of Denmark

*Correspondence:* Charles Boy Kromann, CEKU, Centre for Clinical Education, Rigshospitalet Afsnit 5404, Teilembygningen, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark.  
Tel: 00 45 22 36 88 76; Fax: 00 45 35 45 44 37;  
E-mail: charles.kromann@gmail.com

## INTRODUCTION

It is a general assumption that 'assessment drives learning' through its format, content and programming.<sup>1,2</sup> This is generally believed to pertain to its influence on students' learning strategies.<sup>3</sup> In addition, assessments that include feedback may induce learning from, for example, an objective structured clinical examination (OSCE).<sup>4</sup> However, as well as these *extrinsic* effects of assessment and examinations on students' learning, it has been demonstrated that testing can have an *intrinsic* effect on the memory of studied materials.<sup>5</sup>

In a recent review of studies on the intrinsic effect of testing, Roediger and Karpicke<sup>5</sup> provided evidence that testing students on studied material results in improved retention of that material compared with spending an equivalent amount of time restudying the material. This so-called testing effect has, in both laboratory and classroom studies, been demonstrated to be a robust and independent phenomenon that applies to a variety of test formats and levels of knowledge learning.<sup>5</sup> Wheeler *et al.* described two strengths involved in the act of remembering: storage strength and retrieval strength. Storage is the process induced by study sessions and retrieval is induced by testing.<sup>6</sup> Studies converge on naming repeated retrieval as the key to the testing effect.<sup>5,7,8</sup> The more elaborate the retrieval process is, the better the effect of testing.<sup>9</sup>

Consequently, including testing as part of a course might be an efficient strategy to improve learning outcome. This is especially relevant for courses on topics that cannot be left to individual study, as is the case for a variety of procedural skills. We searched the extensive literature on skills learning and assessment and found that whether the intrinsic testing effect applies to skills learning had never been thoroughly investigated.

Simulation-based skills courses are used to train people in procedures that are difficult to learn in real-life settings.<sup>10</sup> The problem of poor retention of learning outcome following simulation-based courses has been demonstrated repeatedly.<sup>11,12</sup> The risk of decay of learning outcome is especially augmented when opportunities to practise the learned skills after the course are limited, as is the case with in-hospital resuscitation and safe defibrillation. Because of the poor retention of in-hospital resuscitation skills and the unremitting need for health professionals to be able to act appropriately in emergencies, it is

recommended that in-hospital resuscitation courses are repeated frequently.<sup>13,14</sup>

In general, simulation-based skills courses are fairly expensive<sup>15</sup> and are confined to a limited amount of time. Thus, it is important to determine whether learning outcomes are better if some of this time is spent on testing rather than on more training. Although it has been speculated that testing might have only limited effect on skills learning,<sup>16</sup> we found no systematic studies enquiring into whether the impact of the intrinsic testing effect on knowledge retention also pertains to skills learning. A major review by Rosenbaum *et al.* found that 'intellectual and perceptual motor skills are acquired in fundamentally similar ways' in multiple areas like feedback, massed or blocked learning and, most interestingly in view of the area explored in this paper, transfer and retention.<sup>17</sup> A study by Shin and Rosenbaum suggested that motor skills and intellectual skills are co-ordinated in the same manner.<sup>18</sup>

Seen in the light of Roediger and Karpicke's claim – that the testing effect is applicable on all grades and in all settings<sup>5</sup> – these studies lead us to the suggestion that simulation-based teaching might benefit from the integration of testing as a final activity.

The aim of this study was to examine the testing effect related to skills learning in the context of a simulation-based, in-hospital resuscitation skills course for medical students. The research question was: does testing as a final activity, in an in-hospital resuscitation skills course, increase learning outcome compared with an equal amount of time spent practising?

## METHODS

We performed a prospective, controlled, randomised, single-blind, post-test-only intervention study of the testing effect on the learning outcomes of a 4-hour, simulation-based, in-hospital resuscitation skills course.

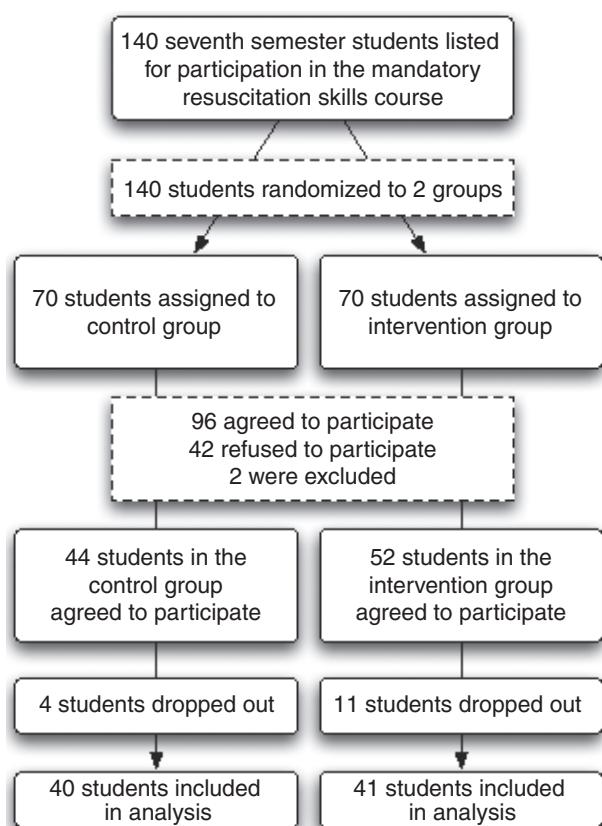
### Study participants

Participants in the study were a volunteer sample of 140 medical students in their seventh semester of study. The students were randomised to receive the intervention course or the control course and were subsequently invited to participate in an assessment of learning outcomes 2 weeks after the course.

Participants were excluded if they had received resuscitation training from other sources within 6 months prior to this study or if they had prepared for the final assessment of learning outcome. Two participants were excluded under the first criterion, but none were excluded under the second. Figure 1 shows a summary of the study profile.

## Settings

The skills course for both the control and intervention groups was a mandatory in-hospital resuscitation course developed for seventh semester medical students. The course covered the essential knowledge and skills needed to start treating adult patients in cardiac arrest. The course was highly standardised in its format and conducted in a skills laboratory associated with the medical school. Students were taught in groups of six. The major part of the course consisted of practical training in resuscitation skills in simulated cardiac arrest scenarios delivered after a short theoretical introduction.



**Figure 1** Profile of the study on the effect of testing on skills learning

## Intervention course

At the beginning of the course, participants in the intervention group were told that their resuscitation skills were to be tested at the end of the course. They were informed that the test was formative in the sense that individuals' results would not have any consequences and the results would not be registered. They were not told that we were investigating a testing effect. Instead, we told them we were comparing different teaching techniques by assessing learning outcomes 2 weeks after the course.

The course designed for the intervention group included 3.5 hours of teaching and training plus 30 minutes of skills testing. After the course itself, the tester, who was new to the group, arrived. The tester was never the same person as the instructor. The tester gave the participants a short briefing. The participants were tested individually in one of six standardised, 5-minute cardiac arrest scenarios, while the others in the group observed. One of these six scenarios was used for the outcome assessment 2 weeks later. The test scenarios differed with regard to the case story, but the scenarios and checklists used for the test were essentially the same and were developed according to European Resuscitation Council guidelines' for shockable and non-shockable algorithms.<sup>19</sup> The tester rated the student's performance on each item on the 25-item checklist during the intervention, using a scale of 0–5. Students were not informed of their individual performance results, but after all six participants had been tested, the tester gave a short group feedback session.

## Control course

The course designed for the control group lasted 4 hours and was identical to the intervention course until the last half-hour. The final 30 minutes were used to run through three to four scenarios, which were identical to those used for the test in the intervention group. One of these scenarios was used for the outcome assessment 2 weeks later. All six participants in each group participated. After each scenario, brief feedback was given to the group.

## Assessment of learning outcome

Assessment of learning outcome was carried out 2 weeks after the course. As the course was offered during two separate rotations spaced with 1 month between them, we scheduled two assessment days offering assessment hours from 0800 h to 2000 h. The participants were tagged with random numbers

from an ID list and instructed not to tell the assessor which type of course they had attended. Each participant was assessed individually without any observers present, except for the assessor. In the assessment room, the assessor ran through a set of standardised instructions for the participants, which included information on utensils, medicines, the degree of assistance available during the scenario and a short case story. The following statement introduced the simulated case: 'You're about to establish an i.v. access when your patient, a 75-year-old man, becomes unresponsive. You are now required to manage this patient.'

Each participant's performance of in-hospital resuscitation skills was assessed using a 25-item checklist. The checklist was inspired by the Advanced Life Support Cardiac Arrest Scenario Test checklist and designed in co-operation with an advanced life support instructor to match the curriculum of the course and expected skill level. Each item was scored on a scale of 0–5, where a score > 2 was regarded as indicating acceptable performance. An item performed with confidence, proper technique and appropriate timing achieved a score of 5. Poor performance or failing to perform the item at all scored 0. One single-blinded assessor performed all assessments. The checklist for assessments is shown in Appendix S1 (available online as supporting information).

### Pilot study

The main study was preceded by a pilot study in order to measure immediate learning outcome, validate the assessment checklist and perform a power calculation. Participants similar to those in the main study group (seventh-semester medical students from the same school) were invited to participate in the pilot study, which used similar designs for the intervention and control groups as described above. A volunteer sample of 40/65 (62%) accepted. The results from the pilot study showed the course to have a large immediate learning effect, with a mean pre-test skills level of 19.7% (95% confidence interval [CI] 15.4–24.0) and a mean post-test skills level of 85.9% (95% CI 83.4–88.4). An assessment of learning outcomes 2 weeks later showed significantly higher learning outcomes in the intervention group ( $n = 22$ ; mean score 79.0%, 95% CI 74.3–83.7) compared with the control group ( $n = 19$ ; mean score 71.5%, 95% CI 66.3–76.7) ( $P = 0.04$ , independent samples  $t$ -test). The effect size (ES) of the difference was 0.65, equivalent to a medium-to-large effect.<sup>20</sup>

### Sample size and statistical analysis

The minimum sample size calculation based on the results from the pilot study showed that 32 subjects in each group would suffice to identify an ES of 0.7 between the group means, with a power of 80% and a significance level of 0.05.<sup>21</sup> The total score from the assessment of learning outcome was converted to a percentage and was compared between the control and intervention groups using independent samples  $t$ -test and reported as a mean (95% CI). The ES estimate was calculated using Cohen's  $d$ , with an ES of 0.2 representing a small ES, 0.5 representing a medium ES, and 0.8 representing a large ES.<sup>20</sup>

All participants gave written informed consent.

---

## RESULTS

In the main study, 81 of 140 (58%) of the invited students chose to participate and completed the trial. Learning outcomes were significantly higher in the intervention group ( $n = 41$ ; mean score 82.8%, 95% CI 79.4–86.2) compared with the control group ( $n = 40$ ; mean score 73.3%, 95% CI 70.5–76.1) ( $P < 0.001$ ). The ES of the difference was 0.93, indicating a large ES.

---

## DISCUSSION

This study suggests that testing as a final activity in a resuscitation skills course increases learning outcome compared with an equal amount of time spent in practice. The mean score of the intervention group was superior to that of the control group, at 82.8% compared with 73.3%, yielding an ES of 0.93.

Bangert-Drowns and Kulik's 1991 review reports that ESs in testing effect studies are generally moderate to high,<sup>22</sup> and Roediger and Karpicke found an ES of 0.82 in knowledge learning in a recent study.<sup>23</sup> Accordingly, our study on skills learning demonstrating an ES of 0.93 corresponds well with prior studies on the testing effect and indicates that the testing effect can be reproduced in skills learning.

### Limitations

Isolating the intrinsic testing effect on skills learning in a classroom setting represents some challenges. In the context of the in-hospital resuscitation course, several extrinsic effects of testing might confound the intrinsic testing effect, including: better preparation

by students; greater familiarity with the test scenarios; observational learning, and intention to learn. We made great effort to eliminate these extrinsic effects of testing at the point of study design.

#### *Preparation by the participants*

As the participants in the intervention group did not know they were to be tested before the start of the course, it is improbable that there was a difference between the groups with regard to the amount of preparation carried out for the course. Moreover, both groups had equal opportunity to prepare for the final assessment 2 weeks later. However, this opportunity was, in reality, minimal. None of them were offered any opportunity to practise on our simulators.

#### *Drop-out rate*

To avoid bias caused by a large assessment time-span, we scheduled only 2 days for the final outcome assessment. On each of these days we assessed 40 participants, from 0800 h to 2000 h. If participants contacted us to cancel, we offered them the opportunity to be assessed on the day before that scheduled. Thus seven of the 81 participants were assessed 1 day before the others. The primary reason for not participating in the assessment was that the students had other appointments (e.g. classes, work etc.) on assessment days. This explains the 40% drop-out rate. We are aware that using volunteer samples results in a positive selection bias (i.e. volunteers often outperform others).<sup>24</sup> However, the two groups in our study were comparable regarding volunteer bias. Furthermore, the ES in our study is of the same magnitude as those of Roediger and Karpicke in studies that did not use volunteer samples.<sup>23</sup> Thus, we find it unlikely that volunteer bias threatens our results.

#### *Participants conferring*

The assessments were completed during 1 day for each of the two rotations in order to limit the extent to which participants conferred with one another. Furthermore, in order to limit the bias arising from conferral among participants, all participants were told about the assessment set-up and content at the time of signing the informed consent. Thus all participants knew that they were to be assessed in a simple cardiac arrest scenario which would include a demonstration of their resuscitation skills and their ability to adhere to the algorithm.

#### *Content familiarity*

The same six basic scenarios were used repeatedly in both courses, ensuring that both groups were thoroughly familiarised with all scenarios. Furthermore, the scenario used for the assessment 2 weeks after the course had been used by both the intervention and control groups as part of the course. Thus we consider both groups to have been equally familiar with the test scenarios.

#### *Observational learning*

The final half-hour of the courses differed with regard to the relative weighting of observational learning and practical involvement. Studies have shown that watching the performance of procedures can have a reinforcing effect on practical training.<sup>25,26</sup> However, the final half-hour included observational learning and practical involvement in both groups and thus we find that if observational learning has influenced our results, it has been to a limited extent.

#### *Intention to learn*

It has been hypothesised that the expectation of later testing stimulates intention to learn, leading to higher learning outcomes.<sup>27</sup> However, a 1984 review by Crocker and Dickinson<sup>28</sup> claims that intentional learning is only beneficial to learning outcome if it prompts rehearsal in the subjects. As our course was very intensive, none of the participants had any opportunity for additional rehearsal and therefore we consider intention to learn to be of limited importance to our results.

Furthermore, participants were excluded if they had prepared for the assessment in the time between the course and the assessment of learning outcome. We consider this policy to have minimised the effect of intention to learn.

#### *Post-test-only design*

Ideally, educational interventional research includes a pre-test. However, in this study examining the effect of testing, pre-testing would have interfered with the intervention and possibly induced an effect on the control group. In our pilot study we used a pre- and post-test design and demonstrated rather low and homogeneous pre-test levels (mean score 19.7%, 95% CI 15.4–24.0). We therefore felt confident in choosing a post-test-only design for the main study in order to prevent bias induced by pre-testing.

## Assessment

All assessments in this study were performed by one assessor using only one test scenario on the assumption that using one assessor only would result in more consistent assessment practice compared with using different assessors. In a recent study, we found high inter-rater and inter-case reliability using checklists to assess resuscitation performance.<sup>29</sup> These high reliabilities probably reflect the narrow and homogeneous content area involved in assessing resuscitation skills and the highly standardised performance indicators in the checklists. The checklists and scenarios for the current study were similar to those used in the prior study<sup>29</sup> and hence we consider that reliability of the assessments did not threaten the results.

## Further research

Although results from this study indicate a testing effect applied to skills learning, additional studies regarding the generalisability of the results are needed. This includes reproducing the testing effect on a wider range and complexity of skills, as well as conducting similar studies in different settings and on a wider range of study populations.<sup>30</sup> Moreover, there is a need for future studies with longer follow-up periods to estimate the effect on long-term retention of learning outcome.

## Educational implications

This study demonstrates that it is quite feasible to implement testing as a final activity in small-group, simulation-based training, even in a rather short programme (i.e. 4 hours). Any provider of medical education seeks to establish the most efficient use of its simulation facilities in order to get the most out of its costly investment. In this context the enhanced learning outcome of testing could prove valuable for all skills laboratories. The testing effect seems to be especially relevant to in-hospital resuscitation training because retention of rarely used skills is known to be poor.<sup>31</sup>

## CONCLUSIONS

We found that testing as a final activity in an in-hospital resuscitation skills course for medical students increased learning outcomes compared with spending an equal amount of time in practice. This indicates that the intrinsic testing effect previously

demonstrated to enhance knowledge retention also pertains to skills learning.

**Contributors:** all authors contributed to the study concept and design. CK carried out the acquisition of data. CK and MJ contributed to the analysis and interpretation of data and to statistical analysis. CK had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of data analysis. All authors contributed to the drafting and critical revision of the manuscript, and approved the final version for publication. CR and MJ supervised the study.

**Acknowledgements:** we wish to thank Deborah Davis for editorial assistance and the following skilled instructors at the Centre for Clinical Education for willingly implementing our study in their daily work: A Kyhnel, A Gustafsson, A B Hasselager, C Bohnstedt, E L Bessmann, H Spielberg, J Sommer, J W Egholm, L R Wahlstén, L I Hennings, M Engdahl, M G Tolsgaard, M Koefoed, M J V Henriksen, M J Andersen, P P Höiby, S V Bjørck, M B Rasmussen, S S Mogensen, I Bostadløkken and M H Larsen. None of these individuals were compensated for their role in the study. We also acknowledge the Centre for Clinical Education, Rigshospitalet, Copenhagen University Hospital, Copenhagen for administrative, technical and material support.

**Funding:** this research was funded by the Centre for Clinical Education.

**Conflicts of interest:** none.

**Ethical approval:** the Biomedical Research Ethics Committee in the Capital Region of Denmark considered that ethical approval was not required for this study.

## REFERENCES

- van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;1:41–67.
- McLachlan JC. The relationship between assessment and learning. *Med Educ* 2006;40:716–7.
- Schoonheim-Klein ME, Habets L, Aartman IHA, van der Vleuten CP, Hoogstraten J, van der Velden U. Implementing an objective structured clinical examination (OSCE) in dental education: effects on students' learning strategies. *Eur J Dent Educ* 2006;10:226–35.
- Brazeau C, Boyd L, Crosson J. Changing an existing OSCE to a teaching tool: the making of a teaching OSCE. *Acad Med* 2002;77:932–6.
- Roediger HL, Karpicke JD. The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci* 2006;1:181–276.
- Wheeler MA, Ewers M, Buonanno JF. Different rates of forgetting following study versus test trials. *Memory* 2003;11:571–80.

- 7 Kuo TM, Hirshman E. Investigations of the testing effect. *Am J Psychol* 1996;109:451–64.
- 8 Karpicke JD, Roediger HL. Repeated retrieval during learning is the key to enhancing later retention. *J Mem Lang* 2007;57:151–62.
- 9 Carpenter SK, DeLosh EL. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Mem Cognit* 2006;34:268–76.
- 10 Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;27:10–28.
- 11 Arthur W Jr. Factors that influence skill decay and retention: a quantitative review and analysis. *Hum Perform* 1998;11:57–102.
- 12 Druckman D, Bjork RA. Optimising longterm retention and transfer. In *The Mind's Eye: Enhancing Human Performance*. Washington: National Academy Press 1991;23–56.
- 13 Berden H, Willems F, Hendrick J, Knape J. How frequently should basic cardiopulmonary resuscitation training be repeated to maintain adequate skills? *BMJ* 1993;306:1576–7.
- 14 Woppard M, Whitfield R, Newcombe RG, Colquhoun M, Vetter N, Chamberlain D. Optimal refresher training intervals for AED and CPR skills: a randomised controlled trial. *Resuscitation* 2006;71:237–47.
- 15 Issenberg SB, McGaghie WC, Hart IR et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999;282:861–6.
- 16 Davies RP, Perkins GD. Recall after cardiac arrest scenario testing. *Resuscitation* 2006;68:239–42.
- 17 Rosenbaum DA, Carlson RA, Gilmore RO. Acquisition of intellectual and perceptual motor skills. *Annu Rev Psychol* 2001;52:453–70.
- 18 Shin JC, Rosenbaum DA. Reaching while calculating: scheduling of cognitive and perceptual motor processes. *J Exp Psychol Gen* 2002;131:206–19.
- 19 Nolan JP, Deakin CD, Soar J, Böttiger BW, Smith G. European Resuscitation Council Guidelines for Resuscitation 2005 Section 4. Adult advanced life support. *Resuscitation* 2005;67:39–86.
- 20 Hojat M, Xu G. A visitor's guide to effect sizes – statistical significance versus practical (clinical) importance of research findings. *Adv Health Sci Educ Theory Pract* 2004;9:241–9.
- 21 Carney PA, Nierenberg DW, Pipas CF, Brooks WB, Stukel TA, Keller AM. Educational epidemiology applying population-based design and analytic approaches to study medical education. *JAMA* 2004;292:1044–50.
- 22 Bangert-Drowns RL, Kulik J. Effects of frequent classroom testing. *J Educ Res* 1991;85 (2):89–99.
- 23 Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves longterm retention. *Psychol Sci* 2006;17:249–55.
- 24 Callahan CA, Hojat M, Gonnella JS. Volunteer bias in medical education research: an empirical study of over three decades of longitudinal data. *Med Educ* 2007;41:746–53.
- 25 Custers E, Regehr G, McCulloch W, Peniston C, Reznick R. The effects of modelling on learning a simple surgical procedure: see one, do one or see many, do one? *Adv Health Sci Educ* 1999;4:123–43.
- 26 Vinter A, Perruchet P. Implicit motor learning through observational training in adults and children. *Mem Cognit* 2002;30:256–61.
- 27 Magill RA. Memory components, forgetting and strategies. *Motor Learning and Control: Concepts and Applications*, 8th edn. New York: McGraw-Hill 2006;221–44.
- 28 Crocker PR, Dickinson J. Incidental psychomotor learning: the effects of number of movements, practice and rehearsal. *J Mot Behav* 1984;16:61–75.
- 29 Ringsted C, Lippert F, Hesselfeldt R, Rasmussen MB, Mogensen SS, Frost T, Jensen ML, Jensen MK, Van der Vleuten C. Assessment of Advanced Life Support competence when combining different test methods—Reliability and validity. *Resuscitation* 2007;75 (1):153–60.
- 30 Wulf G, Shea CH. Principles derived from the study of simple skills do not generalise to complex skill learning. *Psychon Bull Rev* 2002;9:185–211.
- 31 Cooper S, Johnston E, Priscott D. Immediate life support (ILS) training. Impact in a primary care setting? *Resuscitation* 2007;72:92–9.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

#### Appendix S1. Scenario.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than for missing material) should be directed to the corresponding author for the article.

*Received 14 April 2008; editorial comments to authors 3 June 2008; accepted for publication 15 July 2008*